

CMB No. 2704-0188

[illegible]

NSN 750-01-230-5500

Standard Form
Prescribed by GPO
19-102

October 15, 1997

Mr. Harry Koch
ESC/ENS
5 Eglin Street, Building 1704
Hanscom Airforce Base, MA 01731-2116

Dear Mr. Koch:

This letter contains our R & D Status Report covering the period from July 1, 1997 to September 31, 1997 for Contract F19628-95-C-0118, entitled "Applications of the Theory of Distributed and Real-Time Systems to the Development of Large-Scale Timing-Based Systems".

Technical Progress

In the following report, more information about the people mentioned can be found starting from our group's "people" page, at URL <http://theory.lcs.mit.edu/tds/people.html>.

I. Modelling and verification tools

During this quarter, we made considerable progress on the design of a programming language and tools to support our formal approach to system design and analysis. We also worked on two aspects of the foundations, involving advanced features. Specifically:

- We completed a preliminary design of the IOA language for describing distributed systems. The language definition appears in a language manual available. We also outlined the design for several tools to be used with this language, including a simulator, interfaces to a theorem-prover and a model-checker, and a code-generator. M.Eng. student Anna Chefter has begun working on the simulator, and PhD student Mandana Vaziri on an interface to the SPIN model-checker. Some novel aspects of this design are support for programming using levels-of-abstraction (including simulator and theorem-prover support), and code generation based on formal IOA models of externally-provided system services.
- In the course of our work on modelling automated transportation systems (discussed in Section III.B below), we discovered an inadequacy of our preliminary hybrid I/O automaton (HIOA) model, for modelling hybrid (continuous/discrete) systems. Namely, as it stands, the HIOA model cannot express certain types of composition used in some control-theory models, in which effects of discrete actions may be propagated instantaneously through different system components via shared variables. Lynch and summer research visitor Dr. Roberto Segala worked with Prof. Frits Vaandrager of Nijmegen and postdoc Dr. John Lygeros to generalize the model appropriately. Although we have outlined the needed extension, it will involve considerable work to complete it (because it involves re-proving difficult theorems about the composability of liveness properties). Improvements to the presentation are also needed.

19971117 066

- Visiting PhD student Henrik Jensen used the idea of abstraction to a small finite model to verify that the Burns distributed mutual exclusion algorithm satisfies the mutual exclusion property. The general n -process algorithm is mapped to a simple 2-process algorithm whose correctness is shown to imply the correctness of the full algorithm. The correctness implication is based on a general theorem Jensen proved earlier; the needed conditions are proved using the LP theorem-prover (with very little user assistance). The 2-process algorithm is verified using the SPIN model-checker. A paper has been submitted to the TACAS'98 conference.

II. Applications

A. Distributed system building blocks

We continued our work on several building-blocks for fault-tolerant distributed systems. This quarter, we completed our work on "Eventually Serializable Data Services", produced a full TR version of our work on "View-Synchronous Group Communication Services", and made progress in several new directions that build on our view-synchrony work. Specifically:

- The results of M.Eng. Oleg Cheiner's exploratory study and implementation of the ESDS ("Eventually Serializable Data Service") system have been documented in his Masters thesis, completed in September.

Also, a long version of our paper on ESDS has been prepared and submitted for journal publication.

- Fekete, Lynch and Shvartsman's paper on view-synchronous group communication services appeared in PODC'97. This paper contains automaton-based specifications for group communication primitives such as those used in the Isis, Transis, Horus and Psynch systems. In particular, the paper includes specifications for a virtually synchronous group communication (VSGC) service and for a totally ordered broadcast service. Fekete et al. have modelled an algorithm, derived from one of Dolev and his students, that uses VSGC to implement totally ordered broadcast. Also, a full version of the paper, with all proofs, was completed for a TR.

Also, Dr. Myla Archer of NRL has succeeded in verifying more of the invariants this quarter, including the one that we think is the most difficult. (She found and fixed some small errors in our proof while doing this.)

- M.S student Roger Khazan continued his work on modeling a load-balancing replicated data server. His implementation relies on the underlying group-communication service to achieve

fault-tolerance and efficiency. During this reporting period, Khazan completed most of the work associated with the assertional proof of correctness.

Also, short-term visitor Dr. Shlomi Dolev, Segala, and Shvartsman began work on designing an algorithm to load-balance a set of tasks in a partitionable network.

- PhD student Roberto DePrisco, and Fekete, Lynch, and Shvartsman, began consideration of *dynamic quorum* versions of view-synchronous group communication services, in which view changes are restricted so that clients in new views can always obtain information propagated from old views. Requiring that each view's membership comprise a majority of the processes would suffice, since then all pairs of views would have nonempty intersections. However, the service can allow more freedom than this if it learns that the clients in some views have completed exchanges of information. We devised a preliminary version of a dynamic quorum service, together with a preliminary version of a use of our service to implement a totally-ordered broadcast, and a preliminary version of an implementation of our service, using our ordinary view-synchronous group communication service.
- Fekete and Lynch carried out preliminary discussions about two other aspects of view-synchronous group communication: (1) Using a formal notion of "synchronous" service as an intermediate concept in proving correctness of systems that use view-synchronous services. (2) Extending the dynamic quorum idea to "dynamic configurations", in which options of choices of quorums are allowed.

B. Multiprocessor shared memory models

Most of our work in this area during this reporting period involved a new case study on developing the theory needed to understand how to program using weakly coherent memory models. Our results on this case study are still preliminary. We also made new progress on our much-better-developed RAID memory case study. Specifically:

- PhD student Victor Luchangco presented a paper on precedence-based memory models at WDAG'97. A precedence-based memory allows its clients to explicitly specify precedences among operations, which constrain the order in which the operations are to be performed by the memory; this notion generalizes multiple-processor memory models. Within this framework, the paper defines a generalized notion of sequential consistency, as well as a weak consistency requirement called per-location sequential consistency, and characterizes the conditions under which clients will not be able to distinguish the two types of memory. It also proves that the algorithm used by the Cilk system implements a per-location sequentially consistent memory.

- Luchangco and Lynch began formal study of the notion of *release consistency*, introduced by Gharachorloo, et al, and previously studied formally by Gibbons, Merritt, and Gharachorloo. This study, plus discussions with Profs. Leiserson and Arvind, suggested different approaches to memory models, in which (1) correctness conditions are formulated at the client/program boundary (rather than the usual processor/memory boundary), and (2) correctness conditions for programs using weakly coherent memory might not just be the appearance of sequential consistency, but might be application-specific, or might include guarantees of atomicity for several memory operations. This suggests the theory of transactions may be relevant, and we are beginning to explore this avenue.
- Vaziri wrote a draft of a conference paper on her work on modelling and proving correctness of controller algorithms for RAID systems. New work during this reporting period involved notable simplifications over the earlier model and proof.

C. Automated Transportation Systems

Our work this quarter involved finishing a project on vehicle protection systems, continuing our work on analysis and control of automated highway systems (specifically, emergency deceleration of vehicle platoons), and starting a new project on modelling/verification of controlled aircraft systems. All our work is based on our hybrid I/O automaton (HIOA) model for hybrid (continuous/discrete) systems.

- PhD student Carl Livadas completed his work on modelling automated vehicle protection subsystems, as used in the Raytheon Personal Rapid Transit project (PRT 2000). His model allows composition of protectors that rely on each other's correct operation. The correctness proofs of the various protectors are based on the correctness of a generic "abstract protector". Specific cases considered include overspeed protection and collision avoidance, for a variety of track topologies ranging from a straight track to a graph involving multiple Y-shaped merges and diverges. Work this quarter involved finishing the thesis writeup and beginning preparation of a version for conference submission.
- (Automated highway systems:)
Lygeros and Lynch considered the problem of emergency deceleration of a string of vehicles. They sought to establish conditions under which such a maneuver will be safe, in the sense that any collisions are guaranteed to be at low relative velocities. In this quarter, we finalized our model for describing the evolution of this system in the HIOA framework. In the process, we were forced to extend the modeling framework to make it capable of capturing all the phenomena that arise in this problem. We then investigated necessary and sufficient

conditions for safety for a particular deceleration strategy (one where all vehicles brake as hard as possible). Our work is of great importance for the area of AHS, especially for AHS architectures that involve platooning of vehicles.

- (Air-traffic management systems:)

Lygeros and Lynch began considering the problem of verifying the TCAS conflict detection/resolution software. TCAS is a large software system designed to provide pilots with information and resolution advisories about potential threats. We seek to verify that the proposed algorithm guarantees safety, i.e. maintains a minimum separation between the aircraft. In this quarter, we developed a model for the TCAS system, including models for the pilots, aircraft, sensors and the resolution algorithm itself. The model for the algorithm was extracted from the documentation of the TCAS code provided by the TCAS developers. We then developed a framework for proving the safety of the system. We are currently working on the proof.

Our work is important in the area of ATMS because it provides a methodology for formally proving the correctness of protocols before they are implemented. Currently, the protocols are only tested in simulation, a process that does not provide guarantees and is very time consuming.

- In a parallel project, we began the analysis of the Center TRACON automation system (CTAS). CTAS is another large software system designed to provide advisories for air traffic controllers. This project is still at an early stage; we are currently trying to extract a model for the system from the CTAS publications.

D. Communication

- PhD student Mark Smith completed his PhD thesis, entitled "Formal Verification of TCP and T/TCP". The thesis presents a formal abstract specification for TCP/IP transport level protocols, a formal model for TCP, and a proof that TCP satisfies this specification. It also presents a formal description of the experimental protocol T/TCP, a weaker spec for its behavior, and a proof that T/TCP meets the weaker spec. It also shows that T/TCP does not satisfy the stronger TCP specification, and proves an impossibility result saying that the combination of properties ideally desired for T/TCP are not attainable.

E. Probabilistic Systems

Work on probabilistic systems during this quarter was mainly carried out by summer research visitor Segala.

- Segala and Lynch completed work on the long version of their paper modelling and analyzing the Aspnes-Herlihy randomized consensus algorithm. An extended abstract appeared in WDAG'97, as an invited paper in honor of the memory of Anna Pogosyants.
- Segala prepared a paper entitled "A model for randomized distributed computation", which contains the definition of the probabilistic model of his PhD thesis. He is writing a paper on this model for the COMPOS conference.

III. Algorithms and impossibility results

A variety of work on design and analysis of specific fault-tolerant algorithms continued. This quarter, this included work on basic mathematical characterization of fault-tolerant computability, on fault-tolerant concurrent data structures, and on fault-tolerant load-balancing.

- M.S. student Gunnar Hoest and visiting faculty member Prof. Nir Shavit from Tel Aviv University completed their initial work on a mathematical complexity framework for fault-tolerant asynchronous systems. Their work uses topological models and methods to analyze time complexity in the *iterated immediate snapshot* model, a restricted type of atomic snapshot shared memory model. They obtained tight bounds for the approximate agreement problem, and a fundamental time vs. number of names tradeoff for the renaming problem. A paper appeared in PODC'97, and Hoest completed his M.S. thesis.
- Former M.S. student Gio Della Libera and Shavit's work on reactive diffracting trees appeared in SPAA'97. The paper describes a new version of the diffracting tree synchronization primitive that grows and shrinks according to the load on the data structure. They are now writing a full journal version of the paper.
- Shavit and his student Asaph Zemach (from Tel Aviv University) completed work on a highly concurrent priority queue design based on their earlier "combining funnels" data structure. They completed empirical evaluations of the design using the Proteus simulator, and are writing a technical report for conference submission to PODC'98.

Their paper with Upfal of IBM Almaden on the journal version of their SPAA 96 paper providing a mathematical model for analyzing diffracting tree performance was accepted to a special issue of the journal Mathematical Systems Theory. Shavit, Upfal and Zemach presented a PODC'97 paper on a new "wait-free" sorting algorithm – that is, one that will take logarithmic parallel time and will run (though slightly less effectively) even if many processes fail.

- De Prisco and Shvartsman (working with Prof. Bogdan Chlebus of the Institute of Informatics, University of Warsaw, Poland) previously developed a new fault-tolerant algorithm for the "Do-All" problem of performing n tasks using p message-passing processors under the constraint of maintaining message and work efficiency; their paper appeared in WDAG'97. They are currently studying extensions of the DoAll algorithm that cope with unreliable broadcast.

Special Programs and Major Items of Equipment

None.

Changes in Key Personnel

Alex Shvartsman accepted a faculty position at the University of Connecticut. He is also continuing as a research affiliate in TDS.

Oleg Cheiner started his doctoral studies at CMU.

Mark Smith accepted a research position at AT&T Bell Labs.

John Lygeros returns to Berkeley but is continuing his work with TDS.

Trips, Talks and Conferences

1. Nancy Lynch. "Specifying and Using a Partitionable, View-Synchronous Group Communication Service." PODC'97, Santa Barbara, CA, August 1997.
2. Nancy Lynch. "Modelling and Verification of Advanced Vehicle Control Systems Using Techniques from Distributed Computing Theory." University Transportation Center, Cambridge, MA September 18, 1997.
3. Nancy Lynch. "Decomposing Large, Complex, Concurrent Systems into Manageable Building Blocks." AFOSR Annual Review Meeting, Rome, NY September 1997.
4. Alex Shvartsman. "Performing Tasks on Restartable Message-passing Processors." WDAG'97, Saarbrücken, Germany, September 1997.
5. John Lygeros. "Hierarchical Hybrid Control of Large Scale Systems", INRIA, Sophia Antipolis, France, September 22, 1997.
6. John Lygeros. "Hybrid Control and Transportation Applications", VERIMAG, Grenoble, France, September 25, 1997.

7. Roberto DePrisco. "Revisting the Paxos algorithm". *WDAG'97*, Saarbrücken, Germany, September 1997.
8. Victor Luchangco. "Precedence Based Memory Models." *WDAG'97* in Saarbrücken, Germany, September 1997.
9. Roberto Segala. "Verification of the Randomized Consensus Algorithm of Aspnes and Herlihy: a Case Study." *WDAG'97*, Saarbrücken, Germany, September, 1997.
10. Gunnar Hoest. "Towards a topological characterization of asynchronous complexity." *Sixteenth Annual ACM Symposium on Principles of Distributed Computing*, Santa Barbara, CA, August 1997.

Areas of Concern

None.

Statement of Sufficiency

The contractually prescribed effort appears to be sufficient to achieve the objectives of this contract.

Degrees Awarded

Oleg Cheiner. "Implementation and Evaluation of an Eventually-Serializable Data Service." Masters. September 1997.

Gunnar Hoest. "Towards a Topological Characterization of Complexity in Asynchronous, Distributed Systems." Masters. September 1997.

Carolos Livadas, "Formal Verification of Safety-Critical Hybrid Systems," Masters. September 1997.

Mark Smith. "Formal Verification of TCP and T/TCP." PhD. September 1997.

Related Accomplishments

During this reporting period the following papers were submitted for publication, accepted for publication, or published:

Submitted for publication:

- [1] Roberto Segala and Nancy Lynch. A model for randomized distributed computation. Submitted for journal publication.
- [2] Henrik Jensen and Nancy Lynch. A Proof of Burns N-Process Mutual Exclusion Algorithm using Abstraction. Submitted for publication.

- [3] Anna Pogoyants, Roberto Segala, and Nancy Lynch. Verification of the randomized consensus algorithm of Aspnes and Herlihy: a case study. MIT/LCS/TM-555. Submitted for journal publication.
- [4] Nancy Lynch, Nir Shavit, Alex Shvartsman, and Dan Touitou. Timing and Linearizability of Counting Networks. Full Version, 1997. Submitted for journal publication. Short version appeared in PODC'96.
- [5] Alan Fekete, David Gupta, Victor Luchangco, Nancy Lynch and Alex Shvartsman. Eventually-Serializable Data Services. Submitted for journal publication.
- [6] Nancy Lynch. A Three-Level Analysis of a Simple Acceleration Maneuver, with Uncertainties. *AMAST Workshops on Real-Time Systems 95 & 96*, World Scientific Publishing Company, AMAST Series in Computing. Submitted for publication, as book chapter.

Accepted:

- [7] Alan Fekete, M. Frans Kaashoek, and Nancy Lynch. Implementing Sequentially Consistent Shared Objects Using Broadcast and Point-to-Point Communication. *JACM*. To appear.
- [8] John Lygeros and Nancy Lynch. On the Formal Verification of the TCAS Conflict Resolution Algorithms. *36th IEEE Conference on Decision and Control*, San Diego, California, December 10–12, 1997. Extended abstract. To appear.
- [9] John Lygeros, Claire Tomlin and Shankar Sastry. Multi-objective hybrid controller synthesis. *36th IEEE Conference on Decision and Control*, San Diego, California, December 10–12, 1997. To appear.
- [10] George Pappas, Claire Tomlin, John Lygeros, Datta Godbole and Shankar Sastry. A next generation architecture for air traffic management systems. *36th IEEE Conference on Decision and Control*, San Diego, California, December 10–12, 1997. To appear.
- [11] Roberto Segala. Quiescence, Fairness, and the notion of implementation. *Information and Computation*. To appear.
- [12] Roberto Segala, Rainer Gawlick, Jørgen Søgaard-Andersen and Nancy Lynch. Liveness in Timed and Untimed Systems. *Information and Computation*. To appear.

Appearing:

- [13] Alan Fekete, Nancy Lynch, and Alex Shvartsman. Specifying and using a partitionable group communication service. Technical Memo MIT/LCS/TM-570, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, October 1997.

- [14] Nir Shavit and Dan Touitou. Elimination Trees and the Construction of Pools and Stacks. *Theory of Computing Systems (formerly Mathematical Systems Theory)*, 30:645-670, 1997.
- [15] Victor Luchangco. Precedence Based Memory Models. In Marios Mavronicolas and Philippas Tsigas, editors, *Distributed Algorithms 11th International Workshop, WDAG'97*, Saarbrücken, Germany, September 1997 Proceedings, volume 1320 of *Lecture Notes in Computer Science*, pages 111–125, Berlin-Heidelberg, 1997. Springer-Verlag.
- [16] Alan Fekete, Nancy Lynch, and Alex Shvartsman. Specifying and using a partitionable group communication service. In *Proceedings of the Sixteenth Annual ACM Symposium on Principles of Distributed Computing*, pages 53–62, Santa Barbara, CA, August 1997. Long version of this paper exists that includes the proofs. See [13].
- [17] Roberto De Prisco. Technical Report MIT/LCS/TR-717, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, 1997. Masters thesis.
- [18] Bogdan Chlebus, Roberto DePrisco, and Alex Shvartsman. Performing tasks on restartable message-passing processors. In Marios Mavronicolas and Philippas Tsigas, editors, *Distributed Algorithms 11th International Workshop, WDAG'97*, Saarbrücken, Germany, September 1997 Proceedings, volume 1320 of *Lecture Notes in Computer Science*, pages 96–110, Berlin-Heidelberg, 1997. Springer-Verlag.
- [19] Roberto De Prisco, Butler Lampson, and Nancy Lynch. Revisiting the Paxos algorithm. In Marios Mavronicolas and Philippas Tsigas, editors, *Distributed Algorithms 11th International Workshop, WDAG'97*, Saarbrücken, Germany, September 1997 Proceedings, volume 1320 of *Lecture Notes in Computer Science*, pages 111–125, Berlin-Heidelberg, 1997. Springer-Verlag.
- [20] N. Shavit, E. Upfal, and A. Zemach. A wait-free sorting algorithm. In *Proceedings of the Sixteenth Annual ACM Symposium on Principles of Distributed Computing*, pages 121–128, Santa Barbara, CA, August 1997.
- [21] Gunnar Hoest and Nir Shavit. Towards a topological characterization of asynchronous complexity. In *Proceedings of the Sixteenth Annual ACM Symposium on Principles of Distributed Computing*, pages 199–208, Santa Barbara, CA, August 1997.
- [22] Anna Pogosyants, Roberto Segala, and Nancy Lynch. Verification of the randomized consensus algorithm of Aspnes and Herlihy: a case study. In Marios Mavronicolas and Philippas Tsigas, editors, *Distributed Algorithms 11th International Workshop, WDAG'97*, Saarbrücken, Germany, September 1997 Proceedings, volume 1320 of *Lecture Notes in Computer Science*, pages 22–36, Berlin-Heidelberg, 1997. Springer-Verlag.

Papers in progress

John Lygeros and Nancy Lynch. "Strings of Vehicles: Modeling and Safety Conditions."

Oleg Cheiner and Alex Shvartsman. "Implementing and Evaluating an Eventually-Serializable Data Service."

Matteo Frigo and Victor Luchangco. "Computation-Centric Memory Models."

Roberto De Prisco, Alan Fekete, Nancy Lynch, and Alex Shvartsman. "Reasoning about Dynamic Voting Systems."

Mandana Vaziri and Nancy Lynch. "Proving Correctness of a Controller Algorithm for the RAID Level 5 System."

Roberto Segala. "Compositional Verification of Randomized Distributed Algorithms." (For the COMPOS conference.)

Carolos Livadas, Nancy A. Lynch, H. B. Weinberg. "Formal Framework for Modeling and Verifying Safety-Critical Hybrid Systems. "

Stephen J. Garland, Nancy A. Lynch, and Mandana Vaziri, "IOA: a Formal Language for I/O Automata."

Nancy Lynch, Roberto Segala, Frits Vaandrager, and H. B. Weinberg. "Hybrid I/O Automata." Journal version.

Nancy Lynch and Sergio Rajsbaum. "On the Borowsky-Gafni Simulation Algorithm."

Theses in progress

Roger Khazan. "Group Communication as a Base for a Load-Balancing, Replicated Data Service." Masters thesis. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Victor Luchangco. "Consistency Models for Distributed Memories." PhD thesis. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Mandana Vaziri. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Carl Livadas. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Roberto Deprisco. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Kate Dolginova. MEng thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Henrik Jensen. PhD Thesis. "Integration of Deductive and Algorithmic Methods for Verification of Reactive Systems." Aalborg University, Denmark. Visiting MIT.

Awards:

- M.Eng Ekaterina Dolginova was one of two overall winners in the 1997 Computing Research Association (CRA) Outstanding Undergraduates competition, sponsored this year by Microsoft. Also, we just learned that her Spring 1997 UROP position was funded by MIT's Spertus Family fund.
- Segala was invited to talk at the pre-LICS workshop on "Probabilistic Methods in Verification" to be held in Indianapolis in 1998, and was also invited to give a lecture in a session on Hybrid Systems during the conference on "Mathematical Theory of Networks and Systems" that will be held in Padova, Italy, July 6-10, 1998.

Sincerely,

Nancy Lynch (ri)

Nancy Lynch
NEC Professor of Software Science and Engineering

Electrical Engineering and Computer Science
(617)253-7225
lynch@theory.lcs.mit.edu

MIT Laboratory for Computer Science

Applications of the Theory of Distributed Real-Time Systems
To the Development of Large-Scale Timing-Based Systems

Prof. Nancy Lynch, Principal Investigator

R & D Status Report

Program Financial Status

ARPA Contract # F19628-95-C-0118

CLIN # 0002

Quarterly Report (7/97 - 9/97)

Total Base Contract

Current Funding Profile

Equipment

Planned Expenditures	Actual Expenditures at Report Date	% Completion	Budget At Completion	Latest Revised Estimate	Remarks
858,443	493,982	57.54%	858,443	858,443	
574,447	493,982	85.99%		493,982	*
35,308	47,915	135.71%			

* Data reflects all received funding. Current funding is sufficient through 10/97.